

Java-Assisted Internet Document Delivery

Frank L. Walker and George R. Thoma
Lister Hill National Center for Biomedical Communications
National Library of Medicine

Keywords: Java, Internet, HotMed, MedJava, Document Delivery, Email, MIME, Ariel, DocView, World Wide Web, Interlibrary Loan

Abstract: An increasing number of libraries are using the Internet for document delivery, not only for interlibrary loan, but for delivering documents directly to the patron's desktop computer. This has been made possible through the widespread adoption of software such as Ariel and DocView. Ariel, a product of the Research Libraries Group, converts paper-based documents to bitmapped images, and delivers them over the Internet. The National Library of Medicine's DocView, released in January 1998 and now used by 2,000 people in over 70 countries, is primarily designed for library patrons to receive, display and manage documents received from Ariel systems. Despite the relative effectiveness of such technologies, Internet document delivery to the patron's desktop remains a challenge to document delivery librarians for several reasons. First, the librarian may need to assist patrons in acquiring document-viewing software, and in helping them install it on a wide variety of computer platforms. The librarian may also need to help patrons upgrade their viewing software as new versions are released. Because a heterogeneous patron population may be expected to have a diverse range of computer skills, the document delivery librarian may indeed face the challenge of training this population in using the document viewing software. As if this were not enough, the potential problem of copyright abuse always exists when copyrighted material is distributed electronically. How does a library cope with the potential problems introduced with Internet document delivery to the patron's desktop? One potential solution to this dilemma is Java-assisted document delivery. The National Library of Medicine's Lister Hill National Center for Biomedical Communications is developing software code-named HotMed and MedJava, which provide a method of delivering library documents through the World Wide Web by using a Java applet. While HotMed is software that delivers multimedia information over the Internet, MedJava is a Java software applet that HotMed delivers with scanned documents. MedJava enables the recipient to view the documents using a Java-enabled web browser. This paper describes the prototype HotMed and MedJava software, and shows how they can significantly reduce or eliminate the problems described here that libraries face in adopting Internet document delivery to the patron's desktop.

1. Background

Document delivery by libraries and information service providers has evolved over the past two decades. Interlibrary loan has traditionally meant photocopies of journal articles being mailed to other libraries. While photocopies are still mailed to requesters, document delivery libraries have since added facsimile transmission, and more recently,

Internet document delivery. The 1990s decade has seen the arrival of Internet delivery of library documents, especially with the widespread use of the ArielTM system developed and distributed by Research Libraries Group.^{1,2} Ariel has enabled hundreds of libraries to do interlibrary loan electronically via the Internet. It is a technology that is faster than mail, more reliable than fax, and offers higher resolution images than possible through conventional fax. While libraries used Ariel in the first half of this decade for interlibrary loan, the second half has seen more use of the Internet for document delivery to the patron's desktop computer. DocView, a software product developed at the National Library of Medicine, helps librarians achieve the goal of delivering of library documents over the Internet to the patron's desktop.^{3,4}

Running on all WindowsTM operating systems, DocView is software that enables a library patron to receive documents sent by a library's Ariel system. DocView's compatibility with Ariel enables a library or document supplier to use Ariel to scan a printed document and send the resulting images directly to a patron's computer running DocView. The scanning process produces a file of bitmapped images, which are sent via File Transfer Protocol (FTP)⁵ protocol or Multipurpose Mime Email Extensions (MIME) email.⁶ DocView is capable of displaying monochrome bitmapped images in either the Group on Electronic Document Interchange⁷ (GEDI) file format used by Ariel systems, or in the Tagged Image File Format⁸ (TIFF). DocView permits the user to zoom, scroll, pan and rotate document images. A user may "bookmark" pages for easy browsing or printing, and images may be copied for insertion in word processing documents. DocView allows the user to file and organize the received documents through a built-in document management system. Finally, DocView permits the user to forward documents over the Internet to others, using either FTP or MIME email.

An extensive period of beta testing that lasted 2½ years revealed that a large majority of users felt that DocView had improved the delivery of documents from their libraries.⁹ DocView was released in January 1998 and is freely available. Since its release the DocView software has been downloaded by more than 2,000 registered users in more than 70 countries. A web site established to distribute DocView includes an extensive user manual, a report on the DocView beta test, and published papers related to DocView. The software can be downloaded from the DocView home page on this web site:

<http://archive.nlm.nih.gov/proj/docview/project.htm>.

2. The Challenges for Document Delivery Librarians

Despite the promise of technologies such as Ariel and DocView, Internet document delivery to the patron's desktop continues to remain a challenge to many document delivery librarians for several reasons. First there is the issue of acquiring document-viewing software, distributing it and installing it on a potentially large variety of patron computing platforms. If a library is to begin an Internet document delivery service, the document delivery librarian often must bear the burden of informing the library's user community of how to receive and use the digital documents. While many patrons have access to Windows platforms, there are environments where other computing platforms, such as Macintosh and Unix, may be as widely used as Windows. This is the case for

many university campuses. While DocView runs well under Windows, it is not designed to run under any other operating system. For non-Windows platforms, patrons need to seek shareware and commercial alternatives. The librarian often needs to inform its patrons on the alternative document viewing options for all potential platforms. After the patron acquires the software, it needs to be installed on the patron's computer. Once again, the document delivery librarian may need to provide assistance with installation.

In addition to the task of software installation, there is also the potential task of installing new versions of document viewing software as they are released. When delivery vehicles such as Ariel undergo design changes, sometimes these changes require new versions of software for document reception and usage. If patrons desire to use new versions of document viewers, the software needs to be installed on the patron's computer. Once again, the librarian may need to assist patrons.

Because a heterogeneous patron population may be expected to have a diverse range of computer skills, the document delivery librarian may also face the challenge of training this population in using the document viewing software. While the DocView beta test revealed that most users felt that the software was easy to learn to use, a small minority (about five percent) had difficulty. For any new software there will always be a certain percentage of users who need help. The burden for user training again usually falls on the shoulders of the document delivery librarian. The larger and more diverse the patron population, the more time the librarian will need to spend in user training.

As if this was not enough, the potential problem of copyright abuse always exists when copyrighted material is distributed electronically. Most librarians try to ensure that intellectual property rights are observed by asking patrons to adhere to the copyright law. However, once a user receives a document, most document viewing software programs permit some level of functionality that could be used indiscriminately. Examples include document and image editing, retransmission and printing of multiple copies. Some librarians may tend to worry about what happens to an electronically created document after it leaves the library. What will the patron do with it? Will the use be legal or illegal? How does a library cope with the potential problems introduced with Internet document delivery to the patron's desktop?

3. A Potential Solution

A potential solution lies in changing three things: the format of the delivered document, the method of delivery, and the software for document reception / usage. The National Library of Medicine's Lister Hill National Center for Biomedical Communications is developing software code-named HotMed, which provides three methods for delivering scanned library documents through the Internet: FTP, MIME email, and the World Wide Web. In addition to the TIFF (or GEDI) file format delivered by Ariel systems, HotMed can deliver a document using two alternatives: Portable Document Format (PDF) and MedJava. Table 1 lists the options available for document delivery via HotMed.

Document Type	Method of Delivery		
	FTP	MIME Email	WWW
	TIFF	TIFF, PDF	TIFF, PDF, MEDJAVA

Table 1. HotMed Delivery Methods and Document Types

The advantage of using PDF over TIFF as a format for scanned documents is that viewers are freely available for it on three major computing platforms: Windows, Macintosh and UNIX. The Adobe Acrobat Reader™ works uniformly well on all three platforms. On the other hand, the availability of TIFF viewers for the three computing platforms is not consistent. DocView and other TIFF viewers are freely available for the Windows platform, but users of Macintosh and UNIX computers typically need to go to either shareware or commercial vendors to get suitable solutions. The fact that Adobe Acrobat Reader is freely available from one source for all three platforms simplifies the task of a document delivery librarian who needs to equip a patron population with document viewing software.

The primary advantage of MedJava over TIFF and PDF is that it promises to eliminate the requirement for a patron to acquire and install document-viewing software on the computer. All the patron needs is a web browser that is Java-enabled, e.g., capable of running a Java applet within the browser. Practically all computers sold today come equipped with web browsers. Web browsers from Microsoft and Netscape for the three major computing platforms are currently Java-enabled. This capability enables a user of one of these browsers to download a document from a web server running the HotMed software. HotMed delivers a Java applet along with the document to the patron's computer, and the applet renders the document images within the browser's window. The implementation for HotMed is illustrated in Figure 1.

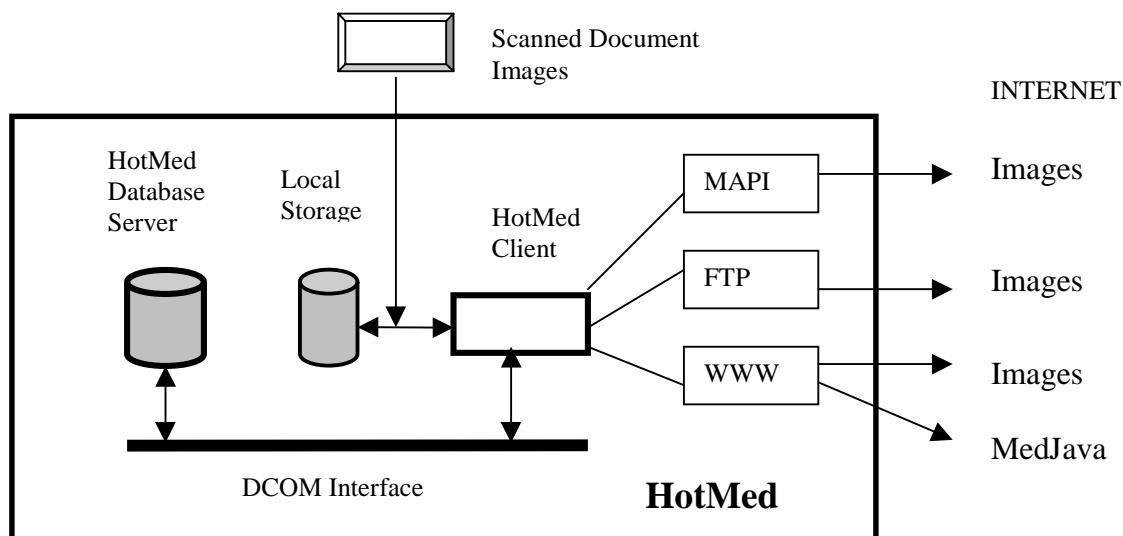


Figure 1. HotMed Configuration

This is a brief description of the components in HotMed:

- Scanned Document Images: Files that will be delivered to the patron.
- HotMed Client: Controls all delivery and provides the user interface.
- HotMed Database Server: A database that keeps track of all deliveries.
- DCOM Interface: A Windows software mechanism called Distributed Component Object Model that allows communication between software components.¹⁰
- Local Storage: A hard disk drive on the computer for temporarily storing document images for delivery.
- MAPI: Windows messaging system called Messaging Application Protocol Interface¹¹ that provides MIME email delivery to patrons. For Web delivery, this provides the method for notifying the patron that the delivery is available on the Web server.
- FTP: Software module for sending documents via the File Transfer Protocol
- WWW: Web server running on the same computer as HotMed.

Document delivery takes place in these steps:

1. The document delivery librarian who is the operator of the HotMed software chooses the scanned document images to be delivered to the patron. The files may be imported from any source, such as scanner, Ariel system, disk or another computer on a local area network.
2. Using an address book, the operator selects the recipient's email address (for MIME email or WWW delivery) or IP address (for FTP delivery).
3. The operator decides how the information is to be delivered: Ariel FTP, MIME email or World Wide Web. Once the method of delivery is chosen, the HotMed client places the delivery information in the HotMed Database Server. The Database Server keeps track of which file is being delivered to which recipient.
4. The HotMed Client delivers the information via the specified mode. FTP transfers the document directly to the recipient's Ariel system or DocView running on their computer. A document delivered via MIME email is automatically converted first to the desired format (TIFF or PDF), then sent as an attachment to the MIME email message to the recipient. The email remains at the patron's email server until he logs in and retrieves the email. If a document is to be delivered via the WWW, the HotMed client software first places the document on the web server running on the HotMed computer. It then notifies the recipient through email of the availability of the document and it provides the patron with the web location (URL) of the document.
5. For web delivery, after a fixed amount of time (nominally 48 hours) after the patron first downloads the document from the Web server, HotMed removes the file from the local storage of the HotMed computer. At the option of the operator, the files may be maintained in local storage for a nominal 21 days if the patron has not retrieved them, and then HotMed will delete them at that point.

The technique for distributing documents using MedJava is slightly more complex. MedJava is a Java applet that HotMed delivers with the document from its web server. Through the patron's web browser, MedJava allows the patron to retrieve, view and print the document located on the web server. MedJava handles monochrome TIFF files,

either uncompressed or compressed. Image expansion algorithms implemented by MedJava include Packbits and ITU recommendations for 1-dimensional Group 3, 2-dimensional Group 3, and Group 4 compression.¹² To help prevent copyright abuse, MedJava does not store the received documents images on the patron's hard disk. Because of this requirement, MedJava itself cannot handle a multipage TIFF file, due to potential memory limitations posed by the host machine. However, prior to delivery to MedJava, HotMed splits the multipage TIFF file into single pages. MedJava is able to retrieve and display one page at a time when sent in this fashion by HotMed.

Usually taking only a few seconds over a high speed Internet connection, there are several additional steps for receiving documents using MedJava, as illustrated in Figure 2.

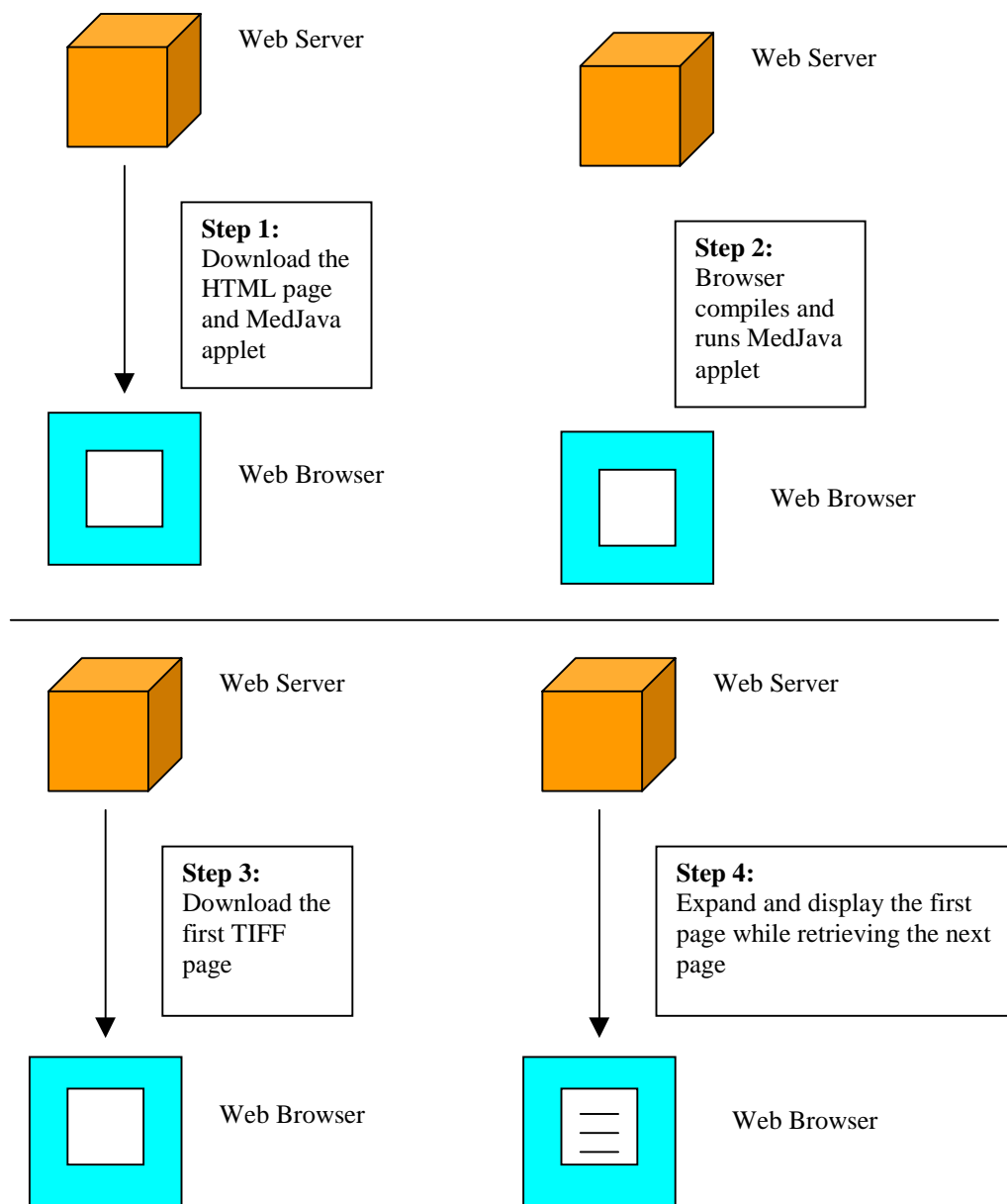
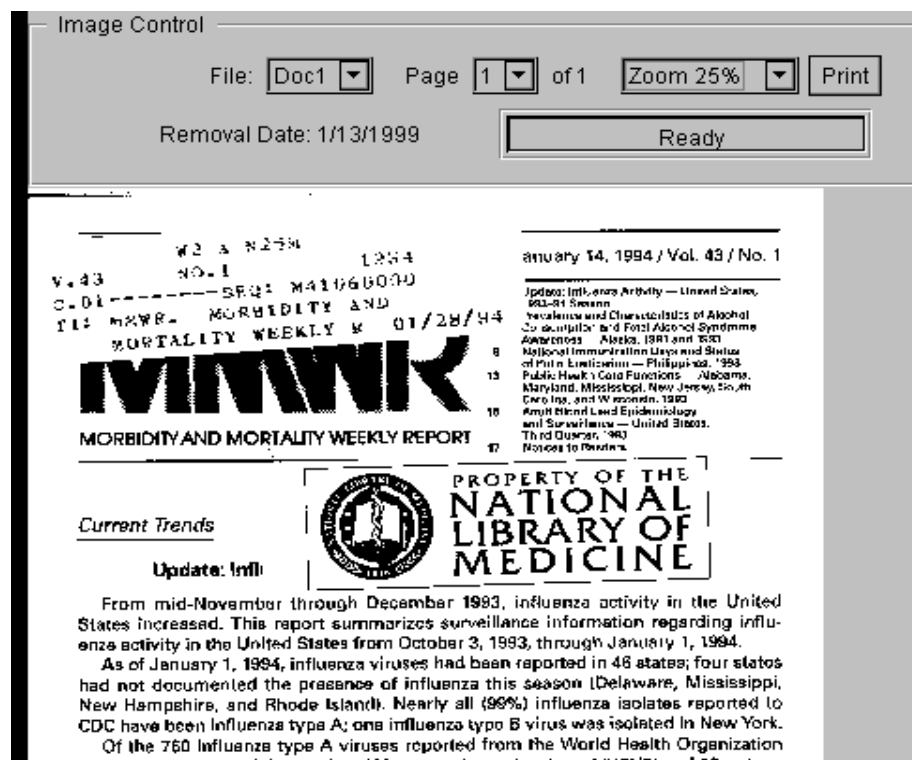


Figure 2. Java-Assisted Document Delivery

These are the four steps for retrieving documents through MedJava:

1. The patron receives the email message from the HotMed computer. This message contains the URL of the web server running on the HotMed computer. Using a mouse, the patron clicks on the URL. Most email client software will automatically run the web browser at this point, and direct the browser to the HotMed web server. The patron's web browser displays the web page from the HotMed web server and automatically downloads the MedJava applet.
2. The Java Virtual Machine in the web browser uses its Just In Time (JIT) Compiler to compile the Java bytecodes in MedJava. Then MedJava begins running on the patron's computer within the context of the web browser.
3. MedJava goes back to the web server and retrieves the first compressed TIFF image file in the document.
4. MedJava expands the compressed data and displays it in the browser. MedJava permits the user to retrieve subsequent pages of the document and print them.

The screen capture in Figure 3 shows that MedJava's user interface is designed to be simple to minimize user need for help. It allows the user to view or print pages in all documents being delivered from the HotMed web server. Other than a zoom, shrink and scroll capability, MedJava has no other user interface features. Unlike DocView, MedJava does not have the ability to store documents on disk, copy them or forward them over the Internet.



Impact of File F **Figure 3. MedJava User Interface**

ery

The file format and delivery technique can have quite an influence on the document delivery librarian. Because document viewers are required for displaying TIFF or PDF files, the librarian needs to make sure that the library's patron population has the correct viewer for each case. The situation for distribution of PDF files makes the job easy for the librarian since PDF viewers are all available for free from one source on the Internet; some cost may be involved in obtaining TIFF viewers. The document delivery librarian may also need to tackle the problem of training patrons with several brands of TIFF viewers. The case is much simpler with PDF, since the use of Adobe Acrobat is similar across all platforms. If user training is necessary, it will be consistent for all platforms for the case of PDF files.

Table 2 compares the three document formats delivered by HotMed, illustrating their impact on issues affecting document delivery librarians.

	TIFF	PDF	MedJava
Same document viewer for all platforms	No	Yes	Yes
Document viewer needs to be acquired and installed on patron computer	Yes	Yes	No
Document viewer software available from one source	No	Yes	Not Applicable
Viewer software may need to be updated as versions change	Yes	Yes	No
Some patrons may need instructional training	Yes	Yes	Not Likely
Same training for all document viewers	No	Yes	Yes
Possible problem protecting intellectual property rights	Yes	Yes	Probably Not

Table 2. Impact of Document Formats upon Librarians

MedJava goes a long way in solving problems that are specific to conventional document viewers. Specifically, it

1. Permits document delivery to multiple computing platforms (Windows, UNIX and Macintosh). Rather than requiring a document viewer to see the received documents, patrons need only have a web browser such as the ones from Microsoft or Netscape. These browsers run on all three computing platforms. Library documents are delivered along with the MedJava applet, which provides image rendering within the context of the browser.
2. Eliminates problems associated with software distribution and installation. Document viewers need to be provided to library patrons and installed on their computers, often a time-consuming process. MedJava solves this problem because it is automatically installed and runs on the patron's computer without user intervention.
3. Eliminates the need for the end user to update the document viewer software. As new versions of document viewers are released, they need to be installed on the patron's

computer. MedJava eliminates problems associated with version control, since the correct version will always be delivered with the library document.

4. Reduces the need for user training and documentation. Document delivery librarians may sometimes need to train users in using document viewers, and provide them written user manuals. MedJava is designed to reduce or eliminate this problem, since its user interface is very simple. Users do not need to learn another software program; they can use their web browser, with which most computer users are already familiar.
5. Solves, to a large extent, the problem of copyright and protection of intellectual material. One problem associated with all document viewers is that an electronic version of the document is sent to the recipient's computer, where it can be stored on hard disk, copied, modified or redistributed. MedJava solves this problem, since it is designed to prevent the recipient from saving the received document on hard disk. Unlike many TIFF and PDF viewers, MedJava does not have a provision that permits copying of the images so that they can be copied, then pasted into word processing programs. MedJava does permit the user to print the received document. It makes no attempt to prevent the user from printing multiple copies of the document.

6.2 Evaluation of Java-Assisted Document Delivery

Both HotMed and MedJava will be extensively evaluated through beta testing. After a period of in-house development and alpha testing, several external organizations will be selected as beta test sites. They will be given the software to use and evaluate for Internet document delivery to their patrons. A user survey will be utilized to determine the overall utility of the software, and its effectiveness in solving the problems that it is designed to solve. The survey will compare Java-assisted document delivery with conventional Internet document delivery. The evaluation of MedJava will address the problems it is intended to solve:

1. *Document delivery to multiple computing platforms.* Is MedJava capable of running well on UNIX, Macintosh and Windows? Are there any browsers on these computers that cause problems for MedJava?
2. *Eliminates problems associated with software distribution and installation.* Is it possible to measure the benefits derived from not having to distribute viewer software and install it on computers of library patrons? It may be necessary to set up a controlled experiment in which a library sends documents to some patrons who will need to install document viewers on their computers, and to other patrons who will use MedJava in place of a document viewer. It may also be necessary to survey both information deliverers and users to determine the total impact of MedJava.
3. *Eliminates the need for the end user to update the document viewer software.* Is it possible to measure the benefit to the end user in not having to update viewer software?
4. *Reduces the need for user training and documentation.* To what extent does this feature benefit the document delivery librarian? How can it be measured?
5. *Solves, to a large extent, the problem of copyright and protection of intellectual material.* To what extent do librarians and their patrons feel that MedJava protects intellectual material? Do further measures need to be taken?

5. Summary

Prototype software is being designed to alleviate some of the problems experienced by document delivery librarians when they deliver documents to patrons over the Internet. Problems arise when delivery is to be made to large heterogeneous patron populations that have a variety of computing platforms. While most libraries currently doing Internet document delivery use Ariel systems to send TIFF files via either FTP or MIME email, this paper considers other alternatives. Delivery of PDF documents, for instance, should simplify the task of document viewer software installation on patron computers. Further improvements occur when delivery is made over the World Wide Web, and Java applet software is used to provide the rendering and use of the documents on the patron's computer. This paper has given an overview of the design of the prototype software code-named HotMed and MedJava, being developed at the R&D labs of the National Library of Medicine. After development this software will be evaluated through beta testing to determine its effectiveness in addressing issues related to document delivery through the Internet.

References

1. Berger, MA. Ariel Document Delivery and the Small Academic Library. College & Undergraduate Libraries, Vol. 3(2). The Haworth Press, 1996; 49-56.
 2. The World Wide Web address for Research Libraries Group is located at this URL: <http://www-rlg.stanford.edu/welcome.html>
 3. Walker FL, Thoma GR. DocView: Providing Access to Printed Literature through the Internet. Proceedings IOLS'95. Medford NJ: Learned Information, 1995; 165-173.
 4. Walker FL, Thoma GR. Internet Document Access and Delivery. Proceedings IOLS'96. Medford NJ: Learned Information, 1996; 107-116.
 5. Postel, J. and Reynolds, J. File Transfer Protocol, Request for Comments #959, October 1985, available through the Internet at URL <http://ds.internic.net/>.
 6. Borenstein, N. and Freed, N. MIME (Multipurpose Internet Mail Extensions), Request for Comments #1341, June 1992, available through the Internet at URL <http://ds.internic.net/>.
 7. Agreements of the Group on Electronic Document Interchange are available through the Internet at URL <http://lib.ua.ac.be/gedi/gediv21.doc>
 8. TIFF Revision 6.0, Aldus Corporation, June 3, 1992.
 9. Walker F.L., Thoma G.R., "Internet Document Delivery: An End User Survey," Proc. IOLS '97. Medford N.J: Information Today, 1997; 145 - 153.
 10. Sessions, R. Com and Dcom : Microsoft's Vision for Distributed Objects, John Wiley & Sons, December 1997.
 11. De La Cruz, I. and Thaler, L. Inside MAPI, Microsoft Press, 1996.
- ITU Publications are available through the Internet at this URL:
<http://www.itu.int/publications/bookstore.html>.